PENGEMBANGAN ARSITEKTUR *BIG DATA* UNTUK *PLOTTING* TREN DAN PEMETAAN DEMAM BERDARAH *DENGUE* DI ASIA TENGGARA MENGGUNAKAN DATA MEDIA SOSIAL *TWITTER*

SKRIPSI

Disusun Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/ Informatika

Disusun oleh:
Irfan Rizqi Prabaswara
24010315130075

DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2019

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Irfan Rizqi Prabaswara

NIM : 24010315130075

Judul : Pengembangan Arsitektur *Big Data* untuk *Plotting* Tren dan Pemetaan Demam Berdarah *Dengue* di Asia Tenggara Menggunakan Data Media Sosial *Twitter*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 10 Oktober 2019

Irfan Rizqi Prabaswara
24010315130075

HALAMAN PENGESAHAN

Judul : Pengembangan Arsitektur *Big Data* untuk *Plotting* Tren dan Pemetaan Demam Berdarah *Dengue* di Asia Tenggara Menggunakan Data Media Sosial *Twitter*

Nama : Irfan Rizqi Prabaswara

NIM : 24010315130075

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **10 Oktober 2019**.

Semarang, 10 Oktober 2019

Mengetahui,

Ketua Departemen Ilmu Komputer/ Informatika

FSM Undip

Dr. Retno Kusumaningrum, S.Si., M.Kom.

NIP. 19810420 200501 2 001

Panitia Penguji Tugas Akhir

Ketua,

Nurdin Bahtiar, S.Si., M.T

NIP. 197907202003121002

HALAMAN PENGESAHAN

Judul : Pengembangan Arsitektur *Big Data* untuk *Plotting* Tren dan Pemetaan Demam Berdarah *Dengue* di Asia Tenggara Menggunakan Data Media Sosial *Twitter*

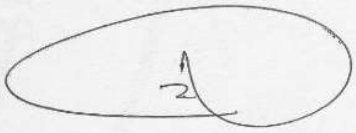
Nama : Irfan Rizqi Prabaswara

NIM : 24010315130075

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **10 Oktober 2019**.

Semarang, 10 Oktober 2019

Dosen Pembimbing



Ragil Saputra, S.Si., M.Cs.
NIP. 198010212005011003

KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya, tugas akhir yang berjudul "Pengembangan Arsitektur *Big Data* untuk *Plotting* Tren dan Pemetaan Demam Berdarah *Dengue* di Asia Tenggara Menggunakan Data Media Sosial *Twitter*" dapat diselesaikan dengan baik. Tulisan ini berisikan dokumentasi mengenai penggunaan *big data twitter* untuk *plotting* tren dan pemetaan demam berdarah *dengue* di Asia Tenggara.

Karya tulis ini dapat diselesaikan dengan baik walaupun banyak sekali kendala dalam menyelesaikan penelitian karena telah dibantu dan dibimbing oleh beberapa pihak. Untuk itu penulis mengucapkan terima kasih kepada :

1. Allah SWT yang sudah meridhoi saya dalam menyelesaikan skripsi ini.
2. Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku Ketua Departemen Ilmu Komputer/Informatika Universitas Diponegoro.
3. Panji Wisnu Wirawan, ST., MT., selaku Koordinator Tugas Akhir.
4. Ragil Saputra, S.Si., M.Cs., selaku Dosen Pembimbing.
5. Nurdin Bahtiar, S.Si., M.T., dan Dr. Eng. Adi Wibowo, S.Si., M.Kom., selaku Dosen Penguji.
6. Keluarga, Majelis Taklim Dolan, Semicolon, Silent Senopati, Uniform, Tim Keju, Tim SBC, Tim Staf Ahli Litbang HMIF, Yuniar Lailatur Rohmah, dan Arif Budiman yang sudah mendukung saya dalam berbagai hal.
7. Seluruh pihak yang telah membantu saya hingga selesainya penyusunan tugas akhir ini yang tidak dapat dituliskan satu per satu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan baik dari segi materi maupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi kesempurnaan skripsi ini. Semoga karya tulis ini dapat bermanfaat bagi pembaca dan untuk pengembangan ilmu pengetahuan khususnya di bidang *big data.*

Semarang, 10 Oktober 2019

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama	: Irfan Rizqi Prabaswara
NIM	: 24010315130075
Program Studi	: Informatika
Departemen	: Ilmu Komputer/Informatika
Fakultas	: Sains dan Matematika
Jenis karya	: Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Noneksklusif (*Non-exclusive RoyaltyFree Right*)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Pengembangan Arsitektur Big Data untuk Plotting Tren dan Pemetaan Demam Berdarah Dengue di Asia Tenggara Menggunakan Data Media Sosial Twitter

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalihmedia/ formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 10 Oktober 2019
Yang menyatakan

Irfan Rizqi Prabaswara

24010315130075

ABSTRAK

Big data merupakan sumber data yang memiliki volume yang besar, variasi yang banyak, dan aliran data yang sangat cepat. Contoh big data antara lain data dari media sosial dan kueri pencarian *Google.* Data tersebut mampu melacak aktivitas penyakit dan data yang ada tersedia setiap saat. Pengolahan *big data* bukanlah suatu hal yang mudah, sehingga diperlukan suatu tools yang dapat membantu proses pengolahan terhadap big data. Salah satu *tools* tersebut adalah *hadoop.* Meskipun kinerja *hadoop* lebih unggul daripada RDBMS tradisional, akan tetapi pengolahan data menggunakan *hadoop* belum maksimal. Sehingga, diperlukan pengolahan data yang lebih cepat. Salah satu cara untuk meningkatkan kecepatan pengolahan data ialah menerapkan *spark* untuk proses pengolahan data yang ada di HDFS. Pada penelitian ini dilakukan *plotting tren* dan pemetaan pada data Demam Berdarah *Dengue* (DBD) yang berasal dari media sosial *twitter.* Penelitian ini bertujuan untuk membuat visualisasi data yang diperoleh dari *twitter* dengan menggunakan *hadoop* dan *spark* dalam memantau perkembangan DBD di wilayah Asia Tenggara. Hasil dari penelitian menunjukkan bahwa *hadoop* dan *spark* dapat digunakan untuk *plotting* tren dan memetakan persebaran DBD. Semakin besar alokasi *memory executor* yang diterapkan serta semakin besar dan serupa alokasi maksimal *memory scheduler* yang diterapkan pada tiap node, maka waktu yang dibutuhkan untuk menyelesaikan *task* semakin singkat. Akan tetapi, pada titik tertentu konfigurasi *hadoop* dan *spark* menemui titik puncaknya, sehingga jika alokasi diperbesar menghasilkan hasil yang sama. Hasil dari *plotting tren* juga menunjukkan adanya hubungan yang kuat antara data *twitter*, data *real* kejadian DBD yang diperoleh dari WHO, dan tren *google.*

Kata Kunci : *Big Data*, *Twitter*, *Plotting* Tren, Demam Berdarah *Dengue*, *Hadoop*, *Spark*

ABSTRACT

Big data is a data source that has a large volume, variations, and very fast data flow. Data from social media and google search queries is an example of big data. Big data can track the disease activities and available to use at any time. Processing big data is complicated, so that it requires a tool to help the process. One of the tools is Hadoop. Although the performance of hadoop is better than traditional RDBMS, the data processing using hadoop haven't reach the maximum level. Because of that, faster data processing is needed. One way to increase the speed is using spark to process the data in HDFS. In this study, conducted trend plotting and mapping on Dengue Hemorrhagic Fever (DHF) data derived from social media twitter. The aims from this study is to create data visualization obtained from twitter using Hadoop and Spark in monitoring of DHF in the Southeast Asia Region. The results shows that Hadoop and Spark can be used to ploting the trends and mapping the distribution of DHF. The greater allocation of memory executor applied as well as the larger and similar maximum allocation of memory scheduler applied to each node, then the time needed to complete the task is shorter. However, at some point configuration of Hadoop and Spark will reach the peak, so that , when the allocation enlarged it will produce the same result. The trends ploting also show a strong relation between twitter data, DHF real data obtained from WHO, and google trends.

Keywords: Big Data, Twitter, Trend Plotting, Dengue Fever, Hadoop, Spark

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI ii
HALAMAN PENGESAHAN iii
HALAMAN PENGESAHAN iv
KATA PENGANTAR v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS vi
ABSTRAK vii
ABSTRACT viii
DAFTAR ISI ix
DAFTAR TABEL xi
DAFTAR GAMBAR xii
DAFTAR LAMPIRAN xiii
BAB I PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 3
1.3. Tujuan dan Manfaat 3
1.3.1. Tujuan 3
1.3.2. Manfaat 3
1.4. Ruang Lingkup 3
1.5. Sistematika Penulisan 4
BAB II LANDASAN TEORI 6
2.1. Tinjauan Pustaka 6
2.2. Dasar Teori 7
2.2.1. Demam Berdarah *Dengue* 7
2.2.2. Pengertian Tren 8
2.2.3. *Twitter* 8
2.2.4. *Big Data* 10
2.2.5. *Apache Hadoop* 11
2.2.6. *Apache Spark* 14
2.2.7. *Spark History Server* 18
2.2.8. *Python* 18
2.2.9. *Plotting* 18

2.2.10. Flowchart 19
BAB III METODOLOGI PENELITIAN 20
3.1. Tahap Awal 21
3.1.1. Identifikasi Masalah 21
3.1.2. Studi Pendahuluan 21
3.2. Tahap Pembangunan Sistem 21
3.2.1. Analisis Kebutuhan Sistem 21
3.2.2. Perancangan Sistem 22
3.3. Tahap Pengumpulan Data 24
3.3.1. Pengumpulan Data *Real* Kejadian Demam Berdarah *Dengue* 24
3.3.2. Pengumpulan Data *Twitter* 24
3.4. Tahap Plotting 27
3.4.1. *Preprocessing Data* 28
3.4.2. Plotting Tren Demam Berdarah *Dengue* 29
3.4.3. Pemetaan Persebaran Demam Berdarah 30
3.4.4. Pengujian dan Penilaian Performa Infrastruktur 32
BAB IV HASIL DAN PEMBAHASAN 34
4.1. Tampilan Program 34
4.2. Hasil Pengumpulan Data 36
4.2.1. Hasil Pengumpulan Data *Real* Kejadian Demam Berdarah *Dengue* 36
4.2.2. Hasil Pengumpulan Data *Twitter* 37
4.3. Hasil Tahap Plotting 38
4.3.1. Hasil Preprocessing Data 38
4.3.2. Plotting Tren di Asia Tenggara 39
4.3.3. Matriks Evaluasi Tren 41
4.3.4. Pemetaan Persebaran Demam Berdarah *Dengue* 41
4.4. Hasil Pengujian Performa Infrastruktur 42
BAB V PENUTUP 46
5.1. Kesimpulan 46
5.2. Saran 46
DAFTAR PUSTAKA 47
LAMPIRAN 50

DAFTAR TABEL

Tabel 2.1. Perbandingan Penelitian Terdahulu 6
Tabel 2.2. Deskripsi Atribut *Tweet* 9
Tabel 2.3. Simbol Flowchart (Subrata, 2019) 19
Tabel 3.1. Spesifikasi Perangkat Keras 22
Tabel 3.2. Konfigurasi Perangkat pada *Cluster Multinode* 23
Tabel 3.3. Konfigurasi Pengujian 33
Tabel 4.1. Data *Real* Jumlah Kejadian DBD Berdasarkan WHO 36
Tabel 4.2. Normalisasi Data *Real* Jumlah Kejadian DBD Berdasarkan WHO 37
Tabel 4.3. Matriks Evaluasi Tren 41
Tabel 4.4. Tabel Hasil Pengujian Alokasi *Memory Executor* 1GB 43
Tabel 4.5. Tabel Hasil Pengujian Alokasi *Memory Executor* 2GB 44
Tabel 4.6. Tabel Hasil Pengujian Alokasi *Memory Executor* 3GB 45

DAFTAR GAMBAR

Gambar 2.1. Siklus Manajemen *Big Data* (Hurwitz, Nugent, Halper, & Kaufman, 2013) 10
Gambar 2.2. Arsitektur *Hadoop* (Apache Software Foundation, 2019) 11
Gambar 2.3. Alur Kerja Proses *MapReduce* (Gunarathne, 2015) 12
Gambar 2.4. Arsitektur HDFS (Ryanto, 2017) 13
Gambar 2.5. Diagram Kerja YARN (Apache Software Foundation, 2018) 14
Gambar 2.6. Komponen *Apache Spark* (Apache Software Foundation, 2019) 15
Gambar 2.7. Arsitektur *Runtime Apache Spark* (Zaharia, Karau, Konwinski, & Wendell, 2015) 15
Gambar 3.1. Garis Besar Metodologi Penelitian 20
Gambar 3.2. Arsitektur *Multinode Cluster* 22
Gambar 3.3. Arsitektur Sistem 23
Gambar 3.4. Diagram Alir Pengambilan Data *Twitter* 25
Gambar 3.5. *Source Code* Mengaktifkan TCP 25
Gambar 3.6. *Source Code Scrapping Tweet* 26
Gambar 3.7. *Source Code String to Spark Dataframe* 26
Gambar 3.8. *Source Code* Menyimpan Data ke HDFS 27
Gambar 3.9. Diagram Alir *Preprocessing Data* 28
Gambar 3.10. Kode Menyatukan *Dataframe* dan Membuat *Alias* 29
Gambar 3.11. Kode *Count Tweet* 29
Gambar 3.12. Diagram Alir Visualisasi Data *Twitter* 30
Gambar 3.13. Kode Mengubah Skema Data 30
Gambar 3.14. Diagram Alir Pemetaan Persebaran DBD 31
Gambar 3.15. Kode *Translate to English* 32
Gambar 3.16. Kode Pemetaan Persebaran DBD 32
Gambar 4.1. Tampilan *Utilities* 34
Gambar 4.2. Tampilan Kelola *Tweet* 35
Gambar 4.3. Penyimpanan Data *Tweet* pada HDFS 37
Gambar 4.4. Data *Tweet* pada Tiap File 38
Gambar 4.5. Hasil Penggabungan Dataframe 38
Gambar 4. 6. Hasil Penghitungan Data 39
Gambar 4.7. Diagram Perbandingan Data Twitter dan Data *Real* WHO 39
Gambar 4.8. Perbandingan Data *Twitter* dan Data *Real* WHO Tiap Negara 40
Gambar 4.9. Peta Persebaran DBD Data *Twitter* Tahun 2017 42
Gambar 4.10. Peta Persebaran DBD Data *Real* WHO Tahun 2017 42
Gambar 4.11. Diagram Hasil Pengujian Alokasi *Memory Executor* 1GB 43
Gambar 4.12. Diagram Hasil Pengujian Alokasi *Memory Executor* 2GB 44
Gambar 4.13. Diagram Hasil Pengujian Alokasi *Memory Executor* 3GB 45

DAFTAR LAMPIRAN

Lampiran 1. Grafik Tren Data *Real*, Data *Twitter*, dan Tren *Google* 51
Lampiran 2. Peta Persebaran Demam Berdarah Berdasarkan Data *Twitter* 59
Lampiran 3. Peta Persebaran Demam Berdarah Menurut WHO 63
Lampiran 4. Hasil Pengujian pada *Spark History Server* 67
Lampiran 5. *Source Code* 78
Lampiran 6. Data *Real* WHO 90

BAB I
PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, manfaat dan tujuan, ruang lingkup, dan sistematika penulisan pada skripsi yang berjudul pengembangan arsitektur *big data* untuk *plotting* tren dan pemetaan demam berdarah *dengue* di Asia Tenggara menggunakan data media sosial *twitter*.

1.1. Latar Belakang

Big data menjadi tren dalam dunia teknologi informasi saat ini. *Big data* merupakan sumber data yang memiliki volume yang besar, variasi yang banyak, dan aliran data yang sangat cepat (Hurwitz, Nugent, Halper, & Kaufman, 2013). Menurut Statistical Analysis System (SAS), *big data* adalah suatu kondisi populer yang digunakan untuk mendefinisikan perkembangan eksponensial serta ketersediaan dari data terstruktur maupun tidak (Basuki, Palit, & Dewi, 2015).

Big data dapat digunakan untuk menggambarkan fenomena yang sedang terjadi saat ini (RA, MJ, & WA, 2014). Contoh *big data* yang dapat digunakan untuk menggambarkan fenomena saat ini adalah data dari media sosial *twitter*. Data tersebut mampu melacak aktivitas demam berdarah *dengue* dan data yang ada tersedia setiap saat (Toledo, et al., 2017). Selain itu, data *twitter* sebagai salah satu big data media sosial juga dapat digunakan untuk mengetahui persebaran *dengue* di Brazil (Carlos, Nogueira, & Machado, 2017).

Pengolahan *big data* bukanlah suatu hal yang mudah (Basuki, Palit, & Dewi, 2015). Pengolahan *big data* tidak dapat disamakan dengan pengolahan data dengan ukuran yang relatif kecil. *Single computer* akan terhambat kinerjanya atau juga tidak akan dapat mengolah data jika ukurannya melebihi kapasitas memori pada komputer tersebut (Ryanto, 2017). Oleh karena itu diperlukanlah suatu *tool* atau kerangka kerja yang dapat membantu proses pengolahan terhadap *big data*.

Salah satu *tools* atau kerangka kerja yang dapat digunakan untuk proses pengolahan *big data* adalah *hadoop*. *Hadoop* merupakan kerangka kerja yang dapat diimplementasikan pada *single computer* ataupun *multiple computer* dalam suatu jaringan tertentu (Basuki, Palit, & Dewi, 2015). *Hadoop* memiliki *MapReduce* sebagai model pemrograman untuk

analisis *big data* dan *hadoop distributed file system* (HDFS) sebagai sistem file yang digunakan untuk menyimpan data yang tidak terstruktur. Selain itu, *hadoop* juga memiliki *yet another resource negotiator* (YARN) yang berfungsi sebagai pengatur *resource* pada seluruh aplikasi di dalam sistem (Apache Software Foundation, 2019).

Penelitian yang dilakukan oleh Ranjan (2017) menyatakan bahwa hadoop lebih *scalable* dan efisien daripada RDBMS tradisional. Selain itu, penelitian Ranjan (2017) yang lain menyatakan bahwa hadoop mempermudah untuk proses pengolahan data dalam waktu yang singkat.

Namun, pengolahan data menggunakan *hadoop* belum maksimal. Saat ini, diperlukan pengolahan data yang lebih cepat untuk memenuhi kebutuhan pengolahan data (Oliviandi, Osmond, & Latuconsina, 2018). Salah satu cara untuk meningkatkan kecepatan dalam pengolahan data ialah menerapkan *spark* untuk proses pengolahan data yang ada di HDFS. *Spark* muncul pada tahun 2012 dengan mengembangkan model *MapReduce* yang ada pada *hadoop* untuk mendukung lebih banyak komputasi secara efektif, seperti *interactive queries* dan *stream processing* (Ryanto, 2017). Kecepatan komputasi yang dilakukan oleh *spark* 100 kali lebih cepat daripada *MapReduce* yang terdapat pada *hadoop* (Apache Software Foundation, 2019).

Penelitian yang dilakukan oleh Ryanto (2017) menyatakan bahwa *spark* menunjukkan kinerja komputasi yang lebih cepat hingga 5 kali lipat daripada *hadoop* pada *cluster* tervirtualisasi. Selain itu, *spark* juga memberikan throughput yang lebih tinggi pada kinerja I/O *cluster* daripada *MapReduce*. Penelitian yang dilakukan oleh Oliviandi (2018) menyatakan bahwa penggunaan spark untuk memproses big data sangatlah tepat karena dapat menurunkan response time rata-rata 50% hingga 70% dari *MapReduce*.

Berdasarkan pemaparan yang telah dijelaskan sebelumnya dapat disimpulkan bahwa data internet sebagai salah satu *big data* dapat digunakan untuk menganalisa, memantau, dan memprediksi terjadinya penyakit. Selain itu, dapat disimpulkan bahwa penggunaan *Hadoop MapReduce*, HDFS, dan Hive lebih efisien dibandingkan dengan penggunaan RDBMS. Namun, permasalahan yang muncul yaitu apakah pengembangan *Hadoop* dan *Spark* secara bersamaan dapat digunakan untuk mengelola *big data* serta apakah pengembangan tersebut menghasilkan performa yang baik dan efisien? Oleh karena itu, pada penelitian ini dilakukan pengembangan *hadoop* dan *spark* untuk *plotting tren* dan pemetaan penyakit. *Plotting tren*

dan pemetaan dilakukan pada kasus demam berdarah *dengue* (DBD) di beberapa negara di Asia Tenggara dengan data yang didapatkan dari media sosial *twitter*.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dirumuskan masalah sebagai berikut :

1. Bagaimana pengembangan hadoop dan spark untuk plotting tren dan pemetaan demam berdarah *dengue*?
2. Bagaimana performa *hadoop* dan *spark* dalam mengelola *big data twitter*?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengembangkan *hadoop* dan *spark* untuk *plotting* tren dan pemetaan demam berdarah *dengue*.
2. Mengetahui performa *hadoop* dan *spark* dalam mengelola *big data twitter*.

1.3.2. Manfaat

Penelitian ini bermanfaat bagi penggiat *big data* dalam mempertimbangkan konfigurasi terbaik yang akan diterapkan pada infrastruktur *big data* yang akan digunakan. Selain itu, hasil dari penelitian ini dapat digunakan sebagai referensi pada penelitian selanjutnya.

1.4. Ruang Lingkup

Ruang lingkup memiliki tujuan memberikan batasan dari skripsi ini agar tidak menyimpang dan sesuai dengan target yang diinginkan. Ruang lingkup dalam skripsi ini adalah sebagai berikut :

1. Data *twitter* yang digunakan ialah data *tweet* dari tujuh negara di Asia Tenggara (Indonesia, Thailand, Singapura, Malaysia, Kamboja, Filipina, dan Brunei Darussalam) pada tahun 2010 sampai 2017.
2. Data *real* yang digunakan ialah data WHO dari tujuh negara di Asia Tenggara (Indonesia, Thailand, Singapura, Malaysia, Kamboja, Filipina, dan Brunei Darussalam) pada tahun 2010 sampai 2017.
3. Infrastruktur *big data* yang digunakan ialah *Apache Hadoop* dan *Apache Spark.*

4. Kata kunci yang digunakan dalam pengambilan data *twitter* yaitu demam berdarah, *aedes*, *dengue fever*, *fogging*, *bệnh sốt xuất huyết dengue*, dan *dengue rashes*.
5. Penelitan ini menggunakan *search analysis* dan tidak menerapkan metode sentimen analisis.

1.5. Sistematika Penulisan

Sistematika penulisan memberikan gambaran laporan dari skripsi ini secara urut dan jelas. Berikut adalah sistematika penulisan skripsi ini :

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, manfaat dan tujuan, ruang lingkup, dan sistematika penulisan dari skripsi ini.

BAB II LANDASAN TEORI

Bab ini membahas teori-teori yang berhubungan dan mendukung topik atau masalah yang dibahas pada skripsi ini seperti tinjauan pustaka, demam berdarah *dengue*, pengertian tren, *twitter*, *big data*, *apache hadoop*, *apache spark*, *spark history server*, *plotting,* dan bahasa pemrograman *python*.

BAB III METODOLOGI PENELITIAN

Bab ini membahas metodologi penelitian yang digunakan dalam menerapkan *apache hadoop* dan *apache spark* untuk memetakan dan menganalisa tren demam berdarah di Asia Tenggara berdasarkan pada data historis *twitter* seperti identifikasi masalah, studi pendahuluan, analisis kebutuhan sistem, perancangan sistem, pengambilan data *real* demam berdarah *dengue*, pengambilan data *twitter*, preprocessing data, *plotting* tren demam berdarah, pemetaan persebaran demam berdarah, serta pengujian dan penilaian performa infrastruktur.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memuat semua temuan ilmiah yang diperoleh sebagai data hasil penelitian. Bagian ini menjelaskan performa apache *hadoop* dan apache *spark* dalam mengelola big data, hasil pemetaan sebaran demam berdarah berdasarkan data historis *twitter* di Asia Tenggara, dan analisa tren demam berdarah berdasarkan data historis *twitter* secara deskriptif.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang dapat diambil dari pengerjaan skripsi ini dan saran-saran yang dapat membantu pada penelitian-penelitian selanjutnya.